Constructing local Cell Specific Networks from Single Cell Data

Xuran Wang[1], David Choi[2] and Kathryn Roeder[1,3,*]

June 2021

[1]Department of Statistics and Data Science, Carnegie Mellon University, Pittsburgh, PA, 15213, USA,

[2] Heinz College, Carnegie Mellon University, Pittsburgh, PA 15213, USA

[3]Computational Biology Department, Carnegie Mellon University, Pittsburgh, PA, 15213, USA

[*] To whom correspondence should be addressed. Email: roeder@andrew.cmu.edu

Abstract

Single-cell RNA sequencing facilitates investigations of gene co-expression networks at the cellular level potentially yielding unique biological insights, but the estimation problem is challenging. We develop a non-parametric approach to estimate cell-specific networks for each cell and cell type and propose novel downstream analyses. The individual networks preserve the cellular heterogeneity and facilitate testing for differences between cell groups. To further our understanding of autism spectrum disorder, we examine the evolution of gene networks in fetal brain cells and compare the gene networks of cells sampled from case and control subjects to reveal patterns in gene co-expression.

Background

Single-cell RNA-sequencing (scRNA-seq) provides a high throughput profile of RNA expression for individual cells that reveals the heterogeneity across cell populations. Recent advances in computational methods enable cell-type classification, novel cell-type identification (39), and trajectory alignment (30); however, among these methods, less attention has been devoted to gene-gene association and transcriptional networks, which can shed light on many vital biological processes. Gene expression is known to be tightly regulated by networks of transcription factors and understanding these networks at a cellular level can identify differences in the inner workings of cells in normal and diseased tissues (10, 40, 28, 11).

Thus far, efforts to estimate co-expression networks from scRNA-seq have been limited in their success (12, 35) for several reasons, including technical challenges such as a sparsity of non-zero counts and a high level of noise (5), nonlinear relationships that are not easily captured by traditional measures, and heterogeneity in co-expression patterns across cells.

Efforts have been made to circumvent these challenges (40, 28, 21, 22, 14). Many approaches aim to estimate a single co-expression network across all cell-types, but to gain access to the subtle differences in gene co-expression at the cellular level we require estimates of gene networks evaluated for each cell-type. Moreover, even if the aim is to estimate a cell-type specific network, traditional methods estimate a single co-expression relationship across the entire sample of cells of that type. With such an approach heterogeneity of co-expression across individual cells is erased, but this knowledge can provide valuable insights for downstream analysis.

To reveal co-expression networks within individual cells, we aim to construct cell-specific networks (CSNs) on the cellular level from scRNA-seq data. This idea was first proposed by Dai et al.(7) and Li et al.(20), wherein they generate individual transcriptional networks at a single-cell level for the first time. The construction of CSN is based on the following assessment of local statistical dependency of a pair of genes: for an individual cell of a particular type, if the co-expression of the pair of genes is unusually high, relative to their distribution assuming the genes to be independent, then we infer a gene-gene relationship (Figure 1). This type of analysis produces a full gene-gene network for each cell that can be envisioned as a graph, with genes as nodes and edges depicting gene-gene dependencies. For each cell, Dai et al.(7) and Li et al.(20) then record the degree sequence of its cell-specific network, to be used as input for clustering or low-dimensional embedding.

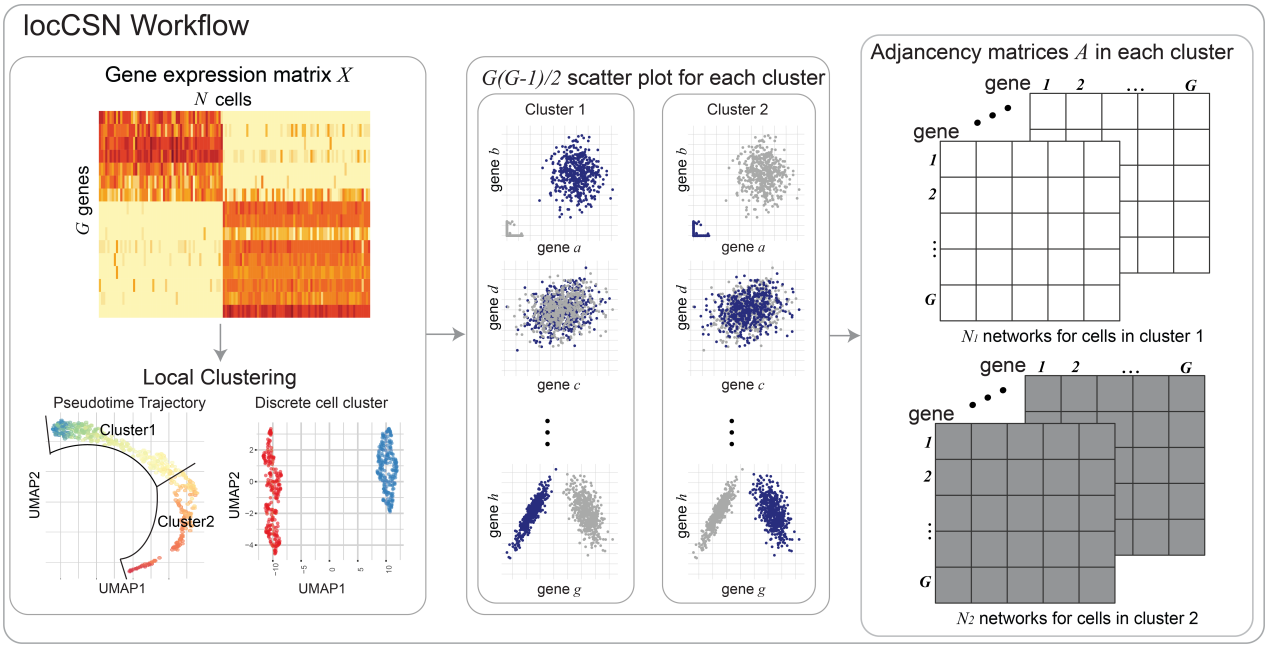


Figure 1: Workflow of locCSN. Starting with the gene expression matrix, we cluster homogeneous cells. For each cell in a cluster, we examine the co-expression of each pair of genes to determine if the joint expression is unusually dense in a neighborhood of the cell, relative to expectation assuming the genes are independent; if so, the genes are connected in that cell's network. For each cell, we then construct the gene-gene network, based on the results of the local independence test for each pair of genes.

We propose an alternative analysis scheme that expands the utility of the CSN concept for gaining insights into cell biology. We retain the individual networks, rather than just their degree sequences, so that differences or developmental changes can be examined at the level of the gene-gene co-expressions. The individual networks can be used in the following ways: First, averaging CSNs across cells within a category or developmental period provides a nonparametric estimator of the gene network that performs well for the sparse counts generally observed for scRNA-seq data, for example allowing for estimation of evolving gene communities in developmental trajectories. Second, having a sample of CSNs provides a valuable measure of the variability across cells that can be leveraged in powerful tests of network differences across cell groups. For example, comparisons can be made between cells obtained from case and control subjects, or cells sampled from different spatial or temporal regimes. Finally, the CSN approach facilitates followup analysis to discover the key genes driving the differences between networks. With this approach we can identify 'differential network (DN) genes', which do not differ in gene expression, but do differ in terms of the co-expression network. These genes could play a key role in network regulation and hence differences that might help explain etiology of disease.

These applications will require more accurate estimates of the individual network connections; we accomplish this by modifying the original network construction algorithm to produce a more adaptive and powerful method, which we call locCSN. We illustrate the value of locCSN by applying it to several scRNA-seq datasets, including developing fetal brains (27), and case and control samples from autism spectrum disorder (ASD) subjects (37). locCSN extracts cell level network information from these data which preserve cell-level network heterogeneity and highlights network differences between case and control samples shedding light on important functional differences.

Results

Our results are summarised as follows. First we describe the locCSN algorithm, Second, simulated and real data examples are used to illustrate differences between locCSN and oCSN. Then, we apply time-varying community detection to CSNs taken from two overlapping developmental trajectories in the developing cortex atlas dataset, and identify differences in their community evolution. Finally, we show that CSNs enable better separation of case and control populations in the ASD brain dataset, by identifying genes that have differing co-expression (but similar expression levels) between the two groups.

CSN construction

The algorithm for construction of CSN as originally published (7) (oCSN) is to estimate for each cell a connection strength between each pair of genes, resulting a gene-gene network for every cell. This connection strength is based on a local independence test that is applied to each gene pair and each cell. To motivate this construction, Dai et al.(7) give examples where the test succeeds in separating a mixture of cells, where genes x and y follow a dependent relationship in some cells but are independent in others (Additional file 1: Figure S2**a**). However, the test requires a choice of resolution (i.e., a bandwidth or window size), which greatly affects performance. Dai et al.(7) proposed a fixed quantile range for the window size, but we find that this approach gives counter-intuitive results and low power when the joint distribution of genes x and y is a simple correlated normal (Additional file 1: Supplementary Note 1 and Figure S1 - S2). To improve performance, we propose a new local algorithm for CSN construction (locCSN) that allows for the window size of the local independence test to vary cell by cell.

Usage

While Dai et al.(7) uses all expressions of all cell-types to construct each cell-specific network, we recommend that each cell-specific network be constructed using only the cells of a common type, so that the independence test is conditional on cell type. Otherwise, false edges may be detected due to Simpson's paradox: given a pair of marker genes that display independent expression within the marked cell-type, CSN construction on all cell-types together will likely infer an edge indicating non-independence of the genes due to the differential expression across types. Therefore we recommend clustering the cells into distinct cell types, and then applying locCSN separately to each cluster, as illustrated for a single cluster in Figure 1. Similarly, if cells are developing along a smooth trajectory, then the cells can be windowed according to pseudotime (30), and each CSN can be computed using the cells within its psuedotime bin.

For each cell, locCSN produces a test statistics for each pair of genes, z_{xy}, to evaluate pairwise gene-gene independence in the neighborhood of that cell's expression. By thresholding these tests we obtain a 0-1 adjacency matrix for each individual cell. Since z_{xy} is approximately normally distributed, a natural threshold is the standard normal quantile, $Z_{(1-\alpha)}$ for α chosen to correspond with a natural choice for significance testing, such as $\alpha = 0.05$ or 0.01. The corresponding entry in the adjacency matrix is $I(z_{xy} > Z_{(1-\alpha)})$, which is either 0 or 1. An estimate of the network for a cluster of cells such as a cell-type is obtained by averaging the adjacency matrices. This average detects unusual clustering of co-expression between genes and thus provides a natural nonparametric estimator of the dependencies in gene expression. The measure is positive by design.

Implementation of locCSN requires two key choices: an initialization of the window width for each local test and the thresholding parameter α to derive the 0-1 adjacency matrix from the matrix of local test statistics. We simulate coexpression networks using ESCO (36) to evaluate performance of network estimation for various choices of these tuning parameters. We observe that the standard deviation approach for window width recommended by locCSN performs markedly better than the quantile approach utilized by oCSN. The performance is robust to choice of starting value for the standard deviation algorithm (Additional file 1: Figure S3 - S6). Our results also show that there is a range of threshold α for which performance is stable and the optimal choice varies somewhat depending on the strength of the true correlations. Nevertheless, for both moderate and strong correlations performance is good between 0.05 and 0.01 after which the accuracy drops precipitously (Additional file 1: Figure S5 - S6).

Illustrative Examples

To illustrate the importance of local calculations we compare locCSN and oCSN for synthetic data from ESCO (36). The cells are sampled from two populations, one for which a set of genes exhibit pairwise correlation and another for which the genes are independent (Additional file 1: Figure S8**a** - **c**). When calculated for this set of genes, ideally the test statistics for each cell will distinguish the pattern of correlation present in one population (Group2) not in the other (Group1) (Additional file 1: Figure S8**c**). The desired pattern is achieved by locCSN, but oCSN produces false signals of correlation for many cells in Group1 (Additional file 1: Figure S8**d** and **e**). Mixing the two populations leads to false positives in Group1 suggesting that the false positives result from Simpson's paradox. Likewise simulations show that BigSCale (14), a powerful alternative approach for coexpression estimation, yields many false connections not detected by locCSN (Additional file 1: Figure S8**f**).

To demonstrate and evaluate the performance of CSN test statistics on real data, we use the Chutype dataset (6). The Chutype dataset includes 7 cell-types (Figure 2**a**). Marker genes corresponding to developmental lineage are provided by the authors and a heatmap of gene expression reveals that a subset of genes

mark cell-types DEC and NPC fairly well (Figure 2**b**). We analyze these two cell-types, which contain 138 cells and 173 cells, respectively. The absolute Pearson's correlation for lineage marker genes, computed within

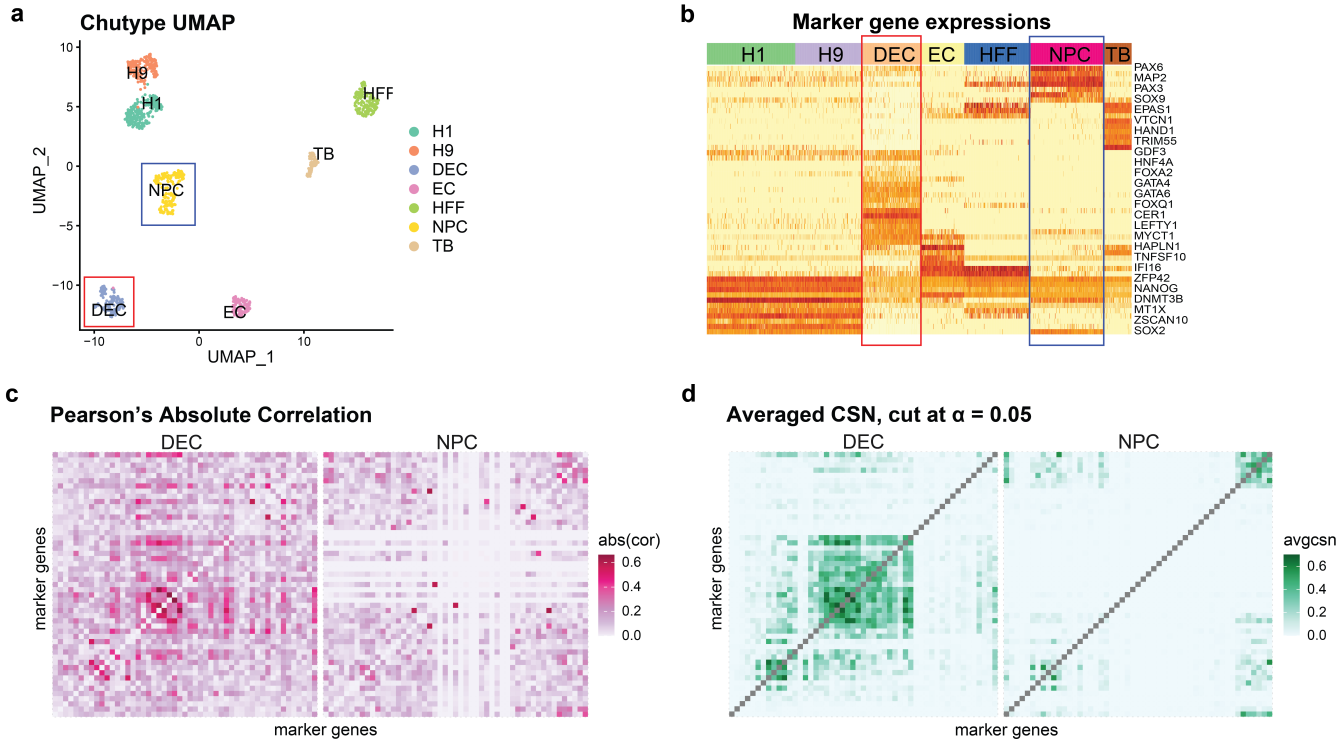


Figure 2: Estimated networks for Chutype cells. (**a**) UMAP of Chutype cells, colored by cell-types. The red and blue boxes indicate selected cell-type: DEC and NPC; (**b**) Heatmap of gene expression for 57 developmental genes for 7 cell-types. High to low expression corresponds to red to light yellow in heatmap. 7 cell-types are color coded by the band on the top of the heatmap. The red and blue boxes indicate DEC and NPC cell-types; (**c**) Heatmaps of absolute Pearson's correlation for DEC and NPC cell-type, calculated independently within cell-types. The order of genes are the same as (**b**); (**d**) Heatmaps of averaged locCSN for DEC and NPC cell-types, thresholded at Gaussian distribution $\alpha = 0.05$ quantile. The order of genes are the same as (**b**).

DEC and NPC cell-types, do not show a clear pattern; specifically, the correlation does not delineate the expected block structure for marker genes for DEC and NPC cells (Figure 2**c**). By contrast, averaged locCSNs, thresholded at $\alpha = 0.05$, preserve the gene block structures and emphasizes the differences between cell-types (Figure 2**d**). We conclude that CSN works well on distinct cell-types. The averaged CSN preserves gene blocks, distinguishes between cell-types, and depicts a clearer co-expression pattern than Pearson's correlation. By contrast, the averaged CSNs calculated from oCSN ignore co-expression between genes, especially the dense block for NPC at the upper right corner (Additional file 1: Figure S9). The performance advantage of locCSN relative to Pearson's is supported by simulations (Additional file 1: Figure S7 and Table S1). These results are notable because the simulation generates a linear relationship between correlated genes. Thus even when the correlation between genes follows the assumed parametric form, locCSN performs better because it adapts to the sparsity expected in scRNA-seq data.

CSN analysis of developmental trajectories

To illustrate the application of CSN to developing cells we feature the developmental trajectory of excitatory neuron cells in the Developing Cortex Atlas dataset, which includes Radial Gila and Progenitors (P), Intermediate Progenitors (IP), Maturing Excitatory (ExN), Migrating Excitatory (ExM), Excitatory Upper-layer enriched (ExM-U), and Excitatory Deep layer (ExDp), with cell-type labels determined by the authors (27) Additional file 1: Supplementary Note 5 and Table S4). Using Slingshot (32), we estimate the developmental path consists of two trajectories, one ending in upper layer (U-curve) and the other in deep layer (D-curve) excitatory neurons (Figure 3**a** and **b**).

To circumvent challenges due to sparse counts, which are prevalent in these data, we pool similar cells within cell-type and form metacells (3), each containing approximately 20 cells. Each metacell's expression,

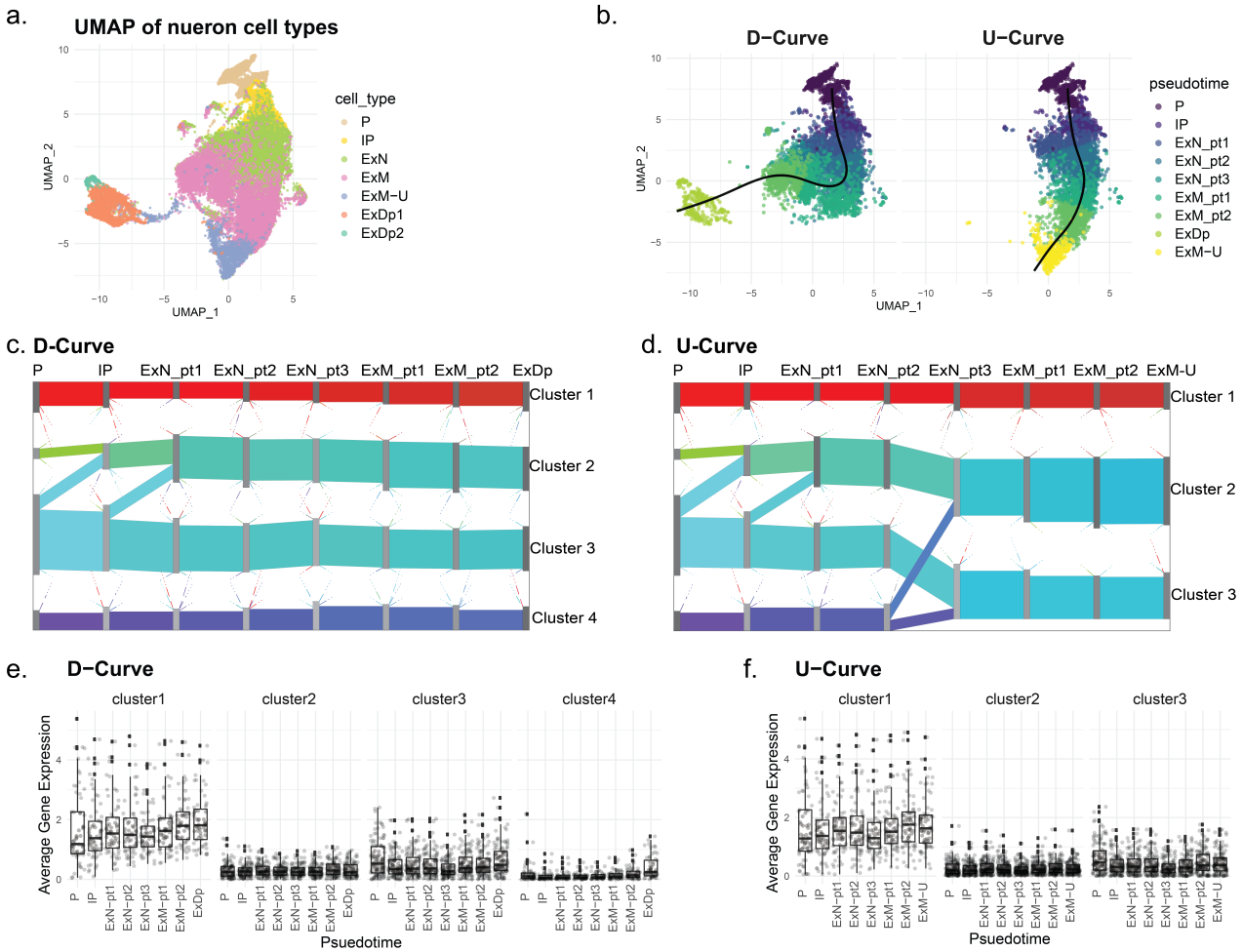


Figure 3: Development of networks in human fetal brain cells. (**a**) UMAP of human fetal brain single-cell expression from 7 cell-types involved in development of excitatory neuron cells, (**b**) with developmental trajectories superimposed. The UMAP plots indicate metacells, whose coordinates are the average UMAP cellular locations. The two principal curves generated by Slingshot (32). Colors are determined by the metacell's pseudotime and cell types for the two curves, which are calculated as the average pseudotime over all cells in the metacell. The left and right panel shows the metacells assignment based on pseudotime to D-curve and U-curve, respectively. (**c,d**) Sankey plots of averaged CSN for 8 bins in D-curve and U-curve. Gene flows are shown as the colored band connecting two adjacent pseudotime bins for (**c**) D-curve and (**d**) U-curve. (**e**) and (**f**) depict boxplots of averaged metacell gene expression for 8 pseudotime bins in the final 4 and 3 clusters for the D- and U-curves, respectively. The x-axis shows pseudotime bins and y-axis shows the averaged expression.

pseudotime and UMAP coordinate is computed as the average over pooled cells. Metacells are assigned to trajectories based on proximity using Slingshot (32). At the early stages of development the curves are nearly overlapping and metacells are not differentiated by curve, but as the cells develop, metacells can be assigned to a distinct curve (Additional file 1: Table S4-S6). Both curves start at the P cell class with full overlap, but moving along ExN and ExM cell classes, there is progressively less overlap between the two curves until they bifurcate, with D-curve and U-curve culminating in purely ExDp and ExM-U metacells, respectively. Next, within each cell-type, we generate bins based on the metacell pseudotime values. This effectively creates bins of approximately 800 metacells containing fairly homogeneous cells.

For each metacell, we use locCSN to compute gene networks. For illustration we focus on a restricted gene list. We choose ASD risk genes because the development of excitatory neurons is of interest in ASD research (27, 25, 31). Specifically CSNs were computed for 444 genes chosen by intersecting the expressed genes in the metacells with a list of 992 ASD-associated genes gleaned from the SFARI database (classes S, 1 and 2) (4). Our objective is to map the formation of gene-clusters over developmental epochs, as cells develop into upper or deep layer excitatory neurons. We apply PisCES (22) to the average CSNs per bin, to find time-varying gene-community structure in the D-curve and U-curve trajectories, and use Sankey plots to visualize the evolution of gene-communities (Additional file 1: Figure 3**c** and **d**). As expected the gene-communities are nearly identical for the two curves for the first four pseudotime bins, which share a large fraction of metacells, but the results progressively differentiate as the overlap in metacells diminishes. Most notably, the gene-community associated with cluster 4 (purple) for U-curve splits and merges into the other gene-communities at ExN_pt3, while it persists for D-curve. For both curves, cluster 1 (red) contains genes that are more highly connected than the other loosely connected gene-communities (Additional file 1: Figure S10). For all of the remaining identified gene-clusters the correlation is extremely weak in the early stages and becomes more apparent in the final developmental bin.

We study the membership in gene-clusters when they are relatively stable without major splitting and merging from ExN_pt3 an onward. We refer to gene-clusters in D-curve as cluster 1-4 and U-curve as cluster 1-3, where labels run from from top to bottom of the display. The numbers of genes in each stable cluster and the overlap between clusters from two curves are shown in the Table 1 (Additional file 2: Table S13). For example, most of the genes from the cluster 1 (dense cluster) for D-curve are included in the corresponding dense cluster, i.e. cluster 1 for U-curve. While the cluster membership is fairly stable along each curve, the strength of correlation changes over time (Additional file 1: Figure S10).

Table 1: Number of genes in gene communities of D-curves and U-curve.

		D-Curve				
		cluster 1	cluster 2	cluster 3	cluster 4	Total
U-Curve	cluster 1	52*	0	5	0	70
	cluster 2	2	73*	9	30	160
	cluster 3	0	9	86*	13	127
	Total	59	103	111	47	

The rows indicate 3 gene clusters from U-curve and the columns show 4 gene clusters from D-curve. The "total" row and column show the total number of genes in each gene cluster. The intersect between U-curve gene clusters and D-curve gene clusters are shown in each row and clusters. The sum of number of overlapped genes does not sum to the total number of genes because some genes are not clustered in both trajectories. * indicates two clusters in D-Curve and U-Curve have a strong overlap with each other.

We also consider the averaged metacell gene expression across pseudotime for each gene-cluster (Figure 3**e** and **f**). Gene expression is relatively stable across cellular development. The dense gene-cluster (cluster 1) has high expressions while the loose gene-clusters have relatively low expression. Yet, even though all genes outside of the dense cluster have lower expression levels, using locCSN, combined with PisCES, we are able to detect subtle correlation and partition genes into gene-communities. It is worth noting that none of these communities are identified by WGCNA (19) (Additional file 1: Figure S11), which relies on Pearson's correlation, but could also be implemented using average CSN matrices for potentially better performance.

To understand the function of gene communities, we check Gene Ontology (GO) terms (2) (Additional file 1: Supplementary Note 7) for the seven gene communities (Additional file 1: Figure S12, Additional file 3, Table S14). For both curves, GO term treemaps for the dense gene community (cluster 1) include metabolic processes, organelle organization and the process of mitosis, which are critical during the fetal stage. For D-curve, cluster 2 is enriched for chromatin organization, which is critical for gene expression regulation. While for U-curve, cluster 2 is enriched for organelle and cellular component organization. The

most loosely structured set of genes, cluster 3 shows no GO enrichment for either curve suggesting it might not be biologically meaningful. Of greatest interest is cluster 4, which is ultimately restricted to D-curve and enriched for synaptic organization, suggesting that these neurons are more mature. This discovery fits with the biological process of neural migration which naturally proliferates the deep layers earlier than the upper layers.

CSN analysis of brain cells from ASD subjects

To demonstrate how CSN is used to contrast network structure in two populations of cells, we analyze the ASD Brain dataset. These data feature single-nuclei RNA-seq (snRNA-seq) data from an ASD study of cortical nuclei assessing 13 cell-types and illustrate our results using excitatory neurons layers: L2/3, L4, L5/6 and L5/6-CC. For details of the full analysis pipeline and results see Additional file 1: Supplementary Note 6 and Figures S13. To illustrate the results, CSNs are computed only for genes implicated in ASD in the literature. We obtain the gene list by intersecting the 992 measured SFARI genes (4) with expressed genes of ASD case and control dataset and 942 genes remain.

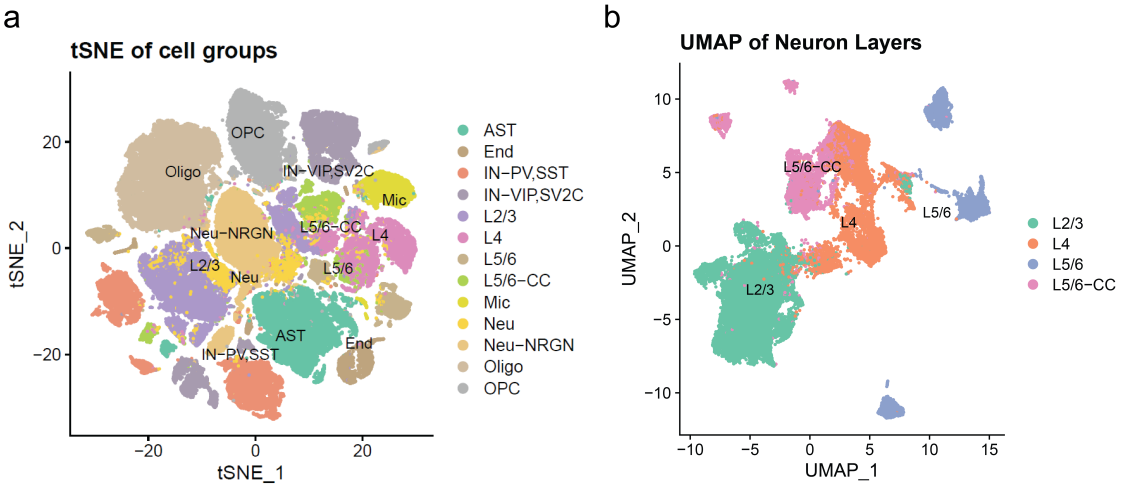


Figure 4: Dimension reduction of brain cells. (**a**) tSNE plot of all brain cells colored by cell-types. (**b**) UMAP plot of Neuron Layers(L) cells colored by 4 cell-types.

After CSN construction, we compare the differences between control and ASD groups by testing for differences in their networks. We use two types of tests: first, an omnibus test for generic differences, and second, a targeted test, aimed at identifying high leverage genes that drive the difference. To test for differences in the distribution of the two classes, we can use a nonparametric distance-based test statistic (DISTp) taken from (23), coupled with a permutation test. This test is highly sensitive to differences in network structure, but a rejection merely indicates that the networks are significantly different. To gain further insights into the network differences, we can utilize the sparse-Leading-Eigenvalue-Driven (sLED) test (41). sLED is designed to detect signal attributable to a modest number of genes in the high dimensional setting encountered for studies of transcription. To emphasize the contrast with differentially expressed genes, we call these differential network genes. sLED takes as input a gene-gene relationship matrix such as the average CSN network or Pearson's correlations matrix. (Further details can be found in Methods.)

The sLED-Pearson test detects no significant differences between control and ASD groups for any cell-type after adjusting for multiple testing, while the sLED-CSN and DISTp tests detect differences in 10 and 6 of the 13 cell-types tested, respectively (Table 2, Additional file 1: Table S10). For each cell-type that yielded a significant sLED-CSN difference we identify the high leverage genes (defined as the top loading genes that explain 90% of the signal by sLED, see Additional file 3: Table S15). We call these the "differential network (DN) genes", to indicate that they strongly differ between case and control in their co-expression and network structure, but not necessarily in expression. The DN genes identified by sLED-CSN show striking differences in their averaged CSNs (Figure 5**a-d**, Additional file 1: Figure S15) and also their visualized networks (Figure 6). Removal of the DN genes reduces the number of detected differences from 10 to 2, suggesting that they explain much of the difference between the groups (Additional file 1: Table S11). Finally, comparing across layers, we find minimal overlap in the DN genes (Figure 7**a**) suggesting that the network differences are layer specific.

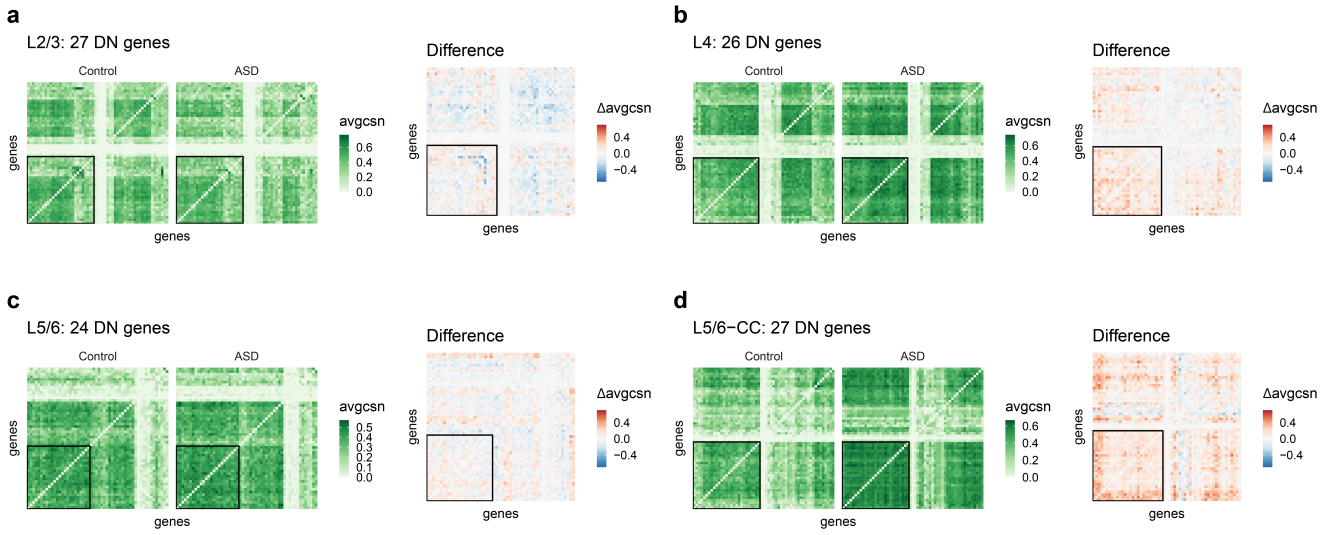


Figure 5: Heatmaps of average CSNs and the average difference (ASD minus Control). Heatmaps display sLED-CSN selected DN genes for each cell-type, contrasted with an additional 30 randomly selected genes from the 942 ASD genes. The black squares delineate the DN genes for each cell-types: (**a**) L2/3; (**b**) L4; (c) L5/6 ;and (**d**) L5/6-CC.

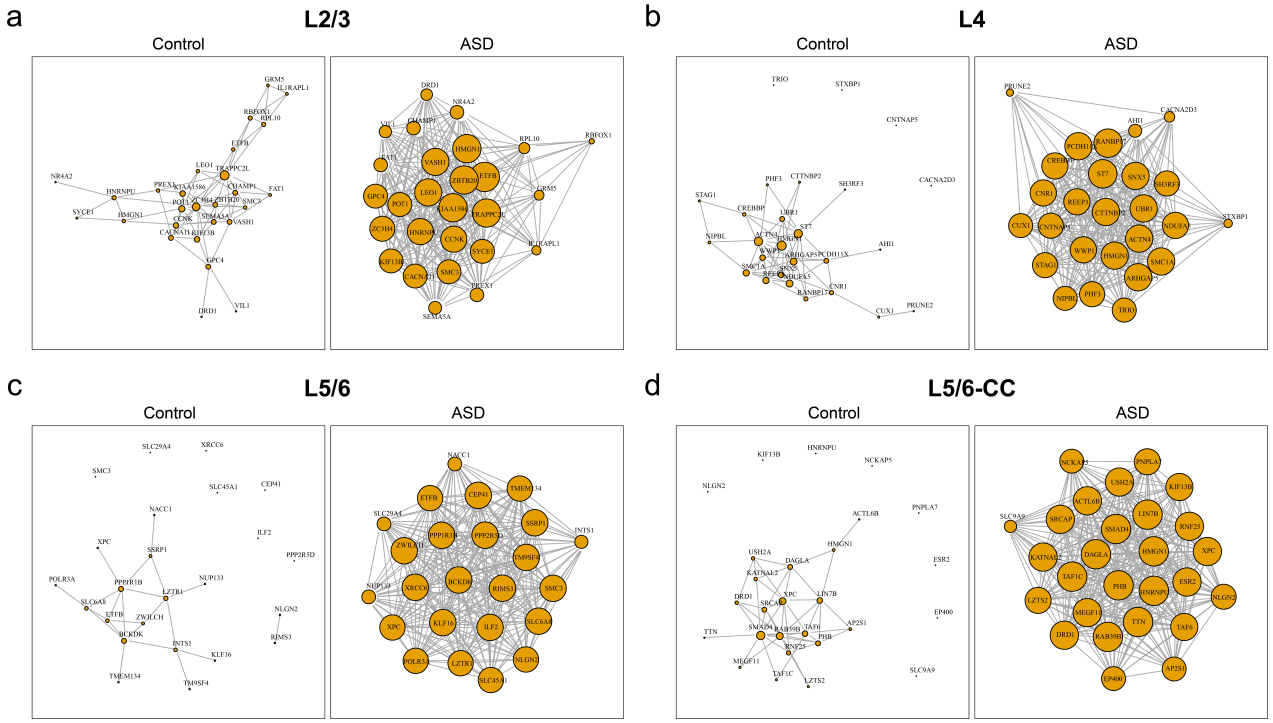
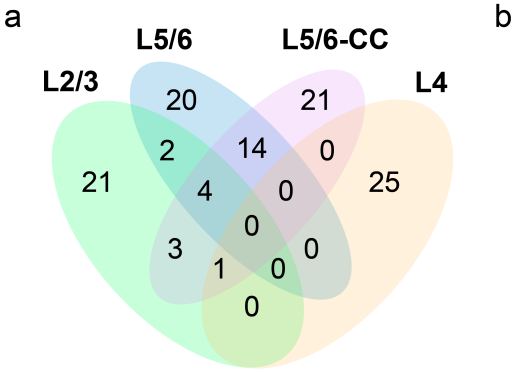


Figure 6: Gene networks for DN genes in the excitatory neuron layers. The networks are generated from averaged CSN of control and ASD groups. (**a**) L2/3; (**b**) L4; (**c**) L5/6; (**d**) L5/6-CC.



b

Cell group	SFARI DE genes	DN genes
AST	12	26
IN-VIP,SV2C	18	26
IN-PV,SST	11	26
L2/3	16	27
L4	17	26
L5/6	1	24
L5/6-CC	2	27
Neu-NRGN	5	26
Oligo	0	28
OPC	1	27

Figure 7: Contrasting differentially expressed and DN genes for ASD vs. Controls. (**a**) Venn diagram of all differential network (DN) genes for neuron layer cell-types. (**b**) Numbers of ASD differentially expressed (DE) genes and DN genes, which have no overlap.

Table 2: P-values for comparing gene expressions and CSNs for control and ASD groups by cell-type.

P-values	L2/3	L4	L5/6	L5/6-CC
sLED-CSN	0.001*	0.001*	0.001*	0.001*
sLED-Pearson	0.089	0.861	0.564	0.348
DISTp	0.001*	0.039	0.001*	0.001*

Comparing gene expressions and CSNs for control and ASD groups, p-values are calculated by sLED-CSN, sLED-Pearson and DISTp for 4 excitatory neuron layers:L2/3, L4, L5/6 and L5/6-CC. * indicates significant differences between two groups after adjusting for multiple testing with 13 cell-types.

For the 10 significant cell-types detected by sLED-CSN, we investigate the overlap between the differential expressed (DE) SFARI genes provided in (37) and the sLED-CSN DN genes. There is surprisingly no overlap (Figure 7**b**), suggesting that most DN genes do not show practically significant differences in expression level between control and ASD groups (Additional file 3: Table S15); similar conclusions can also be drawn by directly inspecting the marginal expression levels (Additional file 1: Figure S16). Overall the DN genes offer critical insights about transcription patterns and how they differ between ASD and control brains.

Methods

locCSN method

To compute $z_{xy}^{(j)}$, the estimated gene-gene relationship of cell j for gene pair (x, y), we first identify the neighborhood and then construct a test statistic. Let $B_x^{(j)}$ and $B_y^{(j)}$ denote one-dimensional bins of cells for genes x and y centered at the expression levels for cell j, with widths w_x and w_y:

$$\begin{aligned} B_x^{(j)} &= \{i : |X_{ix} - X_{jx}| \leq w_x\} \text{ and} \\ B_y^{(j)} &= \{i : |X_{iy} - X_{jy}| \leq w_y\}, \end{aligned}$$

with cell counts $n_x^{(j)} = |B_x^{(j)}|$ and $n_y^{(j)} = |B_y^{(j)}|$. Let $B_{xy}^{(j)} = B_x^{(j)} \cap B_y^{(j)}$ denote the joint window centered at cell j with counts $n_{xy}^{(j)} = |B_{xy}^{(j)}|$.

$z_{xy}^{(j)}$ is represented by a normalized test statistic, $z_{xy}^{(j)} = \rho_{xy}^{(j)}/\sigma_{xy}^{(j)}$, where $\rho_{xy}^{(j)}$ is a local test statistic for independence of genes x and y, comparing the joint distribution with the product of the marginals,

$$\rho_{xy}^{(j)} = \frac{n_{xy}^{(j)}}{N} - \frac{n_x^{(j)} n_y^{(j)}}{N^2},$$

and the normalizing factor $\sigma_{xy}^{(j)}$ is the asymptotic standard deviation of $\rho_{xy}^{(j)}$ under the null hypothesis that genes x and y are independent, shown in (7) to equal

$$\sigma_{xy}^{(j)^2} = \frac{n_x^{(j)} n_y^{(j)} (N - n_x^{(j)})(N - n_y^{(j)})}{N^4 (N-1)}.$$

We remark that as a pre-processing step, we fix $z_{xy}^{(j)} = 0$ if the expression of either gene is zero for that cell.

The choice of the window sizes w_x and w_y plays an important role in the performance of the algorithm. Whereas oCSN by Dai et al. (7) uses window sizes equal to a fixed quantile range, we instead choose the window size based on a local standard deviation, which we compute iteratively. Bins $B_x^{(j)}(0)$ and $B_y^{(j)}(0)$ are initialized to include a quantile range (as in (7)), and then iteratively we follow

$$\begin{aligned} w_x^{(j)}(t) &= \text{St.Dev}\{X_{ix} : i \in B_y^{(j)}(t-1)\} \\ w_y^{(j)}(t) &= \text{St.Dev}\{X_{iy} : i \in B_x^{(j)}(t-1)\} \\ B_x^{(j)}(t) &= \{i : |X_{ix} - X_{jx}| \leq w_x^{(j)}(t)\} \\ B_y^{(j)}(t) &= \{i : |X_{iy} - X_{jy}| \leq w_y^{(j)}(t)\} \end{aligned}$$

for $t = 1, \ldots$ until convergence is achieved (Additional file 1: Figure S2**b**). In practice, if the iterations do not converge, $B_x^{(j)}(1)$ and $B_y^{(j)}(1)$ are used as window sizes.

Metacell: Reduce sparsity of expressions

For single-cell and single-nuclei RNA-seq datasets, the expression can be very sparse. In this setting a direct application of the CSN algorithm fails to discover network structure. It can be advantageous to cluster the data before performing downstream analysis (18), hence, we apply the Metacell algorithm (3) before constructing CSNs. Metacell partitions cells into metacells, defined as disjoint clusters of homogeneous profiles. After applying Metacell to pre-labeled cells, we further divide metacells with multiple cell-types or subtypes into pure-cell-type metacells. Expression of a metacell is defined as the mean of the cells in the cluster, which alleviates the problem of having zero expression for many genes per cell. The metacells are then treated as cells for the purpose of constructing metacell-specific-networks. In this paper, for convenience, CSN refers to either a cell-specific-network or metacell-specific-network.

Trajectory-based community detection

The PiSCES algorithm (22) is designed to identify cluster structure that varies smoothly over adjacent developmental periods. Following CSN construction for a smooth trajectory, where the cells have been binned by similar pseudotimes, we apply the PiSCES algorithm using the average CSN of each bin as input. The time-varying community structure can then be visualized using Sankey plots.

Two sample testing

Given i.i.d. samples of expression levels, the computed CSNs for a set of cells are exchangeable, and hence permutation testing can be used to test for differences in CSN distribution. For this purpose, we suggest two types of tests: first, an omnibus test for generic differences, and second, a targeted test, aimed at identifying high leverage genes that drive the difference.

DISTp: Test CSN differences between groups

Each cell's adjacency matrix can be represented as a vector by converting matrices into vectors, resulting in N_1 sample vectors from class 1 and N_2 sample vectors from class 2. Let $V_1^{(1)}, \ldots, V_{N_1}^{(1)}$ denote the vectorized adjacency matrices from class 1 and $V_1^{(2)}, \ldots, V_{N_2}^{(2)}$ denote the same from class 2. The test statistic Q is a scaled q-norm divergence measurement, with $q \in (0, 2)$ recommended (23), and is given by

$$\begin{aligned} Q(\mathbf{V}^{(1)}, \mathbf{V}^{(2)}; \alpha) &= \frac{2}{N_1 + N_2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} |V_i^{(1)} - V_j^{(2)}|^q \\ &\quad - \frac{N_1 N_2}{N_1 + N_2} \binom{N_1}{2}^{-1} \sum_{1 \leq i < i' \leq N_1} |V_i^{(1)} - V_{i'}^{(1)}|^q \\ &\quad - \frac{N_1 N_2}{N_1 + N_2} \binom{N_2}{2}^{-1} \sum_{1 \leq j < j' \leq N_2} |V_j^{(2)} - V_{j'}^{(2)}|^q, \end{aligned}$$

with p-value calculated by permutation test.

sLED: Identify differential network genes

The sLED test relies on the same principles as Sparse Principal Component Analysis (SPCA), and was originally proposed for the difference in the Pearson's correlation matrices of the two classes (sLED-Pearson). Here we instead propose using the difference in the average CSN as the test input (sLED-CSN). Given N_1 CSN adjacency matrices from class 1 and N_2 from class 2, denoted by $A_1^{(1)}, \ldots, A_{N_1}^{(1)}$ and $A_1^{(2)}, \ldots, A_{N_2}^{(2)}$, let D denote the difference between the average CSN for each class, so that $D = \bar{A}^{(1)} - \bar{A}^{(2)}$. Then D can be used as the input to sLED, in which case the test statistic is computed from the spectrum of D. Additionally, the test also identifies a small cluster of leverage genes corresponding to the non-zero entries of the sparse leading eigenvector. The differential network genes are the ones that explain 90% of the variability among the leverage genes. These are candidate genes that have altered co-expression structure between the two groups. As with DISTp, the p-value of the test statistic is determined by permuting samples among cell classes.

A summary of notation that appears in Results and Method can be found in Additional file 1: Table S2.

Datasets for analysis

Chu-type dataset Chu et al.(6) includes 1018 cells and seven cell-types. This dataset contained the cells of human embryonic stem cell-derived lineage-specific progenitors. The cell-types including H1 embryonic stem cells, H9 embryonic stem cells, human foreskin fibroblasts (HFF), neuronal progenitor cells (NPC), definitive endoderm cells (DEC), endothelial cells (EC) and trophoblast-like cells (TB) were identified by fluorescence-activated cell sorting (FACS) with their respective markers. 9600 genes are obtained per cell on average.

Developing Cortex Atlas dataset Polioudakis et al.(27) includes cells from mid-gestational human cortex. These data are derived from ∼40,000 cells from germinal zones (ventricular zone [VZ], subventricular zone [SVZ]), developing cortex (subplate [SP] and cortical plate [CP]) separated before single cell isolation.

Autism Spectrum Disorder (ASD) Brain dataset Velmeshev et al.(37) includes snRNA-seq data from an ASD study, which collected 105 thousand nuclei from cortical samples taken from 22 ASD and 19 control samples from subjects between 4 and 22 years old. Samples were matched for age, sex, RNA integrity number, and postmortem interval.

Runtime of locCSN

The CSN approach can be computationally intensive because we compute a test statistics for each pair of genes and each cell; however a number of speed ups are possible. By replacing cells with metacells (3), we can reduce the computational complexity substantially. We also provide an approximate CSN calculation by partitioning the outcome space for each pair of genes into a grid. Cells that fall into the same grid yield the same test statistic. With these approximations CSN can be readily applied to very large datasets with good accuracy (Additional file 1: Figure S17). Computational time can also be reduced by limiting the number of genes. While a large number of genes are expressed overall, a relatively modest number of genes are expressed in each cell type and it is natural to restrict analysis to this subset. Another option is to limit investigation to genes of particular biological interest, such as those implicated in disease or involved in biological processes of interest. Alternatively, to construct a network for a large number of genes (≈ 1000), the CSN computation can be applied to each gene pair in parallel.

We observe that, in concordance with the computational complexity, time consumption increases quadratically in the number of genes and linearly in the number of cells or metacell entries. Memory consumption increases in a similar manner. The examples of runtime against input matrix size are shown in Additional file 1: Table S12. Runtimes are measured under Python 3.7.6 [MSC v.1916 32 bit (Intel)].

Discussion

Single-cell gene co-expression networks have the potential to yield critical insights into biological processes. Subtle differences in network structure can be used to classify cells into subtypes and by cell state. While each CSN is estimated with considerable noise, averaging CSNs over homogeneous cells can provide stable estimates of network structure and this can provide insights into how these networks vary by cell type, cell state and over developmental epochs.

Understanding cell-type specific gene networks can contribute substantially to our understanding of how biological processes are impacted by disease and disorders (24, 41, 13, 27). Just as we can test for differential gene expression, we can test for differential coexpression and aim to detect genes that drive the differences in coexpression. Small changes in gene expression can lead to substantial changes in network structure, ultimately with large biological effect. These differential network genes were called 'dark' genes by Dai et al. (7) because we fail to detect them with traditional differential expression and yet can detect them with CSN analysis. For example, we found that most genes that leverage significant differences of connection between ASD and control groups are missed when simply comparing their expression levels. In a similar setting, using the scHOT algorithm, Ghazanfar et al.(11) identified DN genes in developing mouse liver cells that were not DE genes. Both analyses suggest that traditional analyses of gene expression miss critical signals about gene expression differences across developmental epochs and between phenotypes. CSN offers a powerful framework to discover DN genes.

CRISPR-Cas technologies (8, 9, 16) provide researchers with tools to introduce and assess the effects of many genetic perturbations. For example, (17) used the Perturb-Seq method (9) to introduce dozens of ASD risk genes to developing mice brains and then assessed the impact on the single-cell transcriptome. Because CSN estimates the network of each cell it is a natural tool for analysis of such perturbation experiments, which target individual cells with distinct perturbations. CSN analysis provides a useful tool for determining how such perturbations impact the network structure of a cell.

Constructing gene networks using the CSN approach is computationally intensive because we compute a test statistics for each pair of genes and each cell; however, a number of approximations can be readily applied to enhance the speed so that CSN can be applied to very large datasets. It is often natural to reduce the genes under investigation by CSN to a meaningful subset, such as genes previously implicated in genetic risk, genes mapped to critical pathways, or highly variable genes. Restricting the investigation to a subset of genes greatly reduces the computational complexity of CSN analysis, but more importantly, it can reveal more scientifically interpretable results. By focusing on hundreds of documented ASD risk genes, we were able to identify intriguing network structure in developmental trajectories and changes in network structure between ASD and control subjects. In the literature several papers have implicated particular cell-types, especially neurons and subtypes of neurons, in ASD risk (1, 26, 27, 31, 37, 38), but no consensus has been reached. Here we found that gene network structure differed subtly between the developmental trajectories of fetal brain cells for upper layer and deep layer excitatory neurons. Specifically, while both trajectories revealed clustering of genes involved in gene expression regulation, only the deep layer trajectory showed clustering of synaptic genes. Identifying differences in gene networks, both over developmental epochs and between phenotypes can shed light on the genetic etiology of human phenotypes.

Conclusions

locCSN is a refined approach for generating co-expression networks at the level of the individual cell by sharing strength at the level of the gene pair. The output of locCSN can be used to uncover differences in co-expression between cells, and as examples we have shown how changes in co-expression may occur over a developmental trajectory, and also between case and control samples from autism spectrum disorder subjects. Our results highlight the importance of gene co-expression in addition to well-studied gene expression and provide a new perspective for gene modules detection and gene regulation in biological processes. We have made locCSN available at https://github.com/xuranw/locCSN.

Data availability

All data used in this manuscript is publicly available. The accession numbers and links of third-party high-throughput sequencing data obtained from the GEO and EBI database were listed in Additional file 1, Tables S3, respectively.

Software

All of the code used in these analyses is covered by the MIT License and is available from GitHub: https://github.com/xuranw/locCSN.

Acknowledgements

The authors are grateful to Kevin Lin and Jinjin Tian for helpful comments.

Funding

This work was supported in part by National Institute of Mental Health (NIMH) grants R01MH123184 and R37MH057881.

Authors' contributions

This study was conceived of and led by K.R., D.C. and X.W. X.W. designed the model and estimation algorithm, implemented the locCSN software, and led the data analysis. K.R. and D.C. provided scientific insight on downstream analysis methods and data interpretation. X.W., K.R., and D.C. wrote the original draft. The authors read and approved the final manuscript.

Conflict of interest statement.

None declared.

References

[1] Anahita Amiri, Gianfilippo Coppola, Soraya Scuderi, Feinan Wu, Tanmoy Roychowdhury, Fuchen Liu, Sirisha Pochareddy, Yurae Shin, Alexias Safi, Lingyun Song, Ying Zhu, André M M Sousa, PsychENCODE Consortium, Mark Gerstein, Gregory E Crawford, Nenad Sestan, Alexej Abyzov, and Flora M Vaccarino. Transcriptome and epigenome landscape of human cortical development modeled in organoids. *Science*, 362(6420), 12 2018.

[2] Michael Ashburner, Catherine A Ball, Judith A Blake, David Botstein, Heather Butler, J Michael Cherry, Allan P Davis, Kara Dolinski, Selina S Dwight, Janan T Eppig, et al. Gene ontology: tool for the unification of biology. *Nature genetics*, 25(1):25–29, 2000.

[3] Yael Baran, Akhiad Bercovich, Arnau Sebe-Pedros, Yaniv Lubling, Amir Giladi, Elad Chomsky, Zohar Meir, Michael Hoichman, Aviezer Lifshitz, and Amos Tanay. Metacell: analysis of single-cell rna-seq data using k-nn graph partitions. *Genome biology*, 20(1):1–19, 2019.

[4] Saumyendra N Basu, Ravi Kollu, and Sharmila Banerjee-Basu. Autdb: a gene reference resource for autism research. *Nucleic Acids Res*, 37(Database issue):D832–6, Jan 2009.

[5] Shuonan Chen and Jessica C Mar. Evaluating methods of inferring gene regulatory networks highlights their lack of performance for single cell gene expression data. *BMC Bioinformatics*, 19(1):232, 06 2018.

[6] Li-Fang Chu, Ning Leng, Jue Zhang, Zhonggang Hou, Daniel Mamott, David T Vereide, Jeea Choi, Christina Kendziorski, Ron Stewart, and James A Thomson. Single-cell rna-seq reveals novel regulators of human embryonic stem cell differentiation to definitive endoderm. *Genome biology*, 17(1):173, 2016.

[7] Hao Dai, Lin Li, Tao Zeng, and Luonan Chen. Cell-specific network constructed by single-cell rna sequencing data. *Nucleic acids research*, 47(11):e62–e62, 2019.

[8] Paul Datlinger, André F Rendeiro, Christian Schmidl, Thomas Krausgruber, Peter Traxler, Johanna Klughammer, Linda C Schuster, Amelie Kuchler, Donat Alpar, and Christoph Bock. Pooled crispr screening with single-cell transcriptome readout. *Nat Methods*, 14(3):297–301, 03 2017.

[9] Atray Dixit, Oren Parnas, Biyu Li, Jenny Chen, Charles P Fulco, Livnat Jerby-Arnon, Nemanja D Marjanovic, Danielle Dionne, Tyler Burks, Raktima Raychowdhury, Britt Adamson, Thomas M Norman, Eric S Lander, Jonathan S Weissman, Nir Friedman, and Aviv Regev. Perturb-seq: Dissecting molecular circuits with scalable single-cell rna profiling of pooled genetic screens. *Cell*, 167(7):1853–1866.e17, Dec 2016.

[10] Frank Emmert-Streib, Matthias Dehmer, and Benjamin Haibe-Kains. Gene regulatory networks and their applications: understanding biological and medical problems in terms of networks. *Front Cell Dev Biol*, 2:38, 2014.

[11] Shila Ghazanfar, Yingxin Lin, Xianbin Su, David Ming Lin, Ellis Patrick, Ze-Guang Han, John C Marioni, and Jean Yee Hwa Yang. Investigating higher-order interactions in single-cell data with schot. *Nat Methods*, 17(8):799–806, 08 2020.

[12] Minzhe Guo, Hui Wang, S Steven Potter, Jeffrey A Whitsett, and Yan Xu. Sincera: a pipeline for single-cell rna-seq profiling analysis. *PLoS computational biology*, 11(11):e1004575, 2015.

[13] Mads E Hauberg, John F Fullard, Lingxue Zhu, Ariella T Cohain, Claudia Giambartolomei, Ruth Misir, Sarah Reach, Jessica S Johnson, Minghui Wang, Manuel Mattheisen, Anders Dupont Børglum, Bin Zhang, Solveig K Sieberts, Mette A Peters, Enrico Domenici, Eric E Schadt, Bernie Devlin, Pamela Sklar, Kathryn Roeder, Panos Roussos, and CommonMind Consortium. Differential activity of transcribed enhancers in the prefrontal cortex of 537 cases with schizophrenia and controls. *Mol Psychiatry*, 24(11):1685–1695, 11 2019.

[14] Giovanni Iacono, Ramon Massoni-Badosa, and Holger Heyn. Single-cell transcriptomics unveils gene regulatory network plasticity. *Genome Biol*, 20(1):110, 06 2019.

[15] Giovanni Iacono, Elisabetta Mereu, Amy Guillaumet-Adkins, Roser Corominas, Ivon Cuscó, Gustavo Rodríguez-Esteban, Marta Gut, Luis Alberto Pérez-Jurado, Ivo Gut, and Holger Heyn. bigscale: an analytical framework for big-scale single-cell data. *Genome research*, 28(6):878–890, 2018.

[16] Diego Adhemar Jaitin, Assaf Weiner, Ido Yofe, David Lara-Astiaso, Hadas Keren-Shaul, Eyal David, Tomer Meir Salame, Amos Tanay, Alexander van Oudenaarden, and Ido Amit. Dissecting immune circuits by linking crispr-pooled screens with single-cell rna-seq. *Cell*, 167(7):1883–1896.e15, Dec 2016.

[17] Xin Jin, Sean K Simmons, Amy Guo, Ashwin S Shetty, Michelle Ko, Lan Nguyen, Vahbiz Jokhi, Elise Robinson, Paul Oyler, Nathan Curry, Giulio Deangeli, Simona Lodato, Joshua Z Levin, Aviv Regev, Feng Zhang, and Paola Arlotta. In vivo perturb-seq reveals neuronal and glial abnormalities associated with autism risk genes. *Science*, 370(6520), 11 2020.

[18] Tae Hyun Kim, Xiang Zhou, and Mengjie Chen. Demystifying "drop-outs" in single-cell umi data. *Genome Biol*, 21(1):196, 08 2020.

[19] Peter Langfelder and Steve Horvath. Wgcna: an r package for weighted correlation network analysis. *BMC bioinformatics*, 9(1):559, 2008.

[20] Lin Li, Hao Dai, Zhaoyuan Fang, and Luonan Chen. Ccsn: Single cell rna sequencing data analysis by conditional cell-specific network. *bioRxiv*, 2020.

[21] Chee Yee Lim, Huange Wang, Steven Woodhouse, Nir Piterman, Lorenz Wernisch, Jasmin Fisher, and Berthold Göttgens. Btr: training asynchronous boolean models using single-cell expression data. *BMC Bioinformatics*, 17(1):355, Sep 2016.

[22] Xiaoping Liu, Yuetong Wang, Hongbin Ji, Kazuyuki Aihara, and Luonan Chen. Personalized characterization of diseases using sample-specific networks. *Nucleic Acids Res*, 44(22):e164, 12 2016.

[23] David S Matteson and Nicholas A James. A nonparametric approach for multiple change point analysis of multivariate data. *Journal of the American Statistical Association*, 109(505):334–345, 2014.

[24] Aivi T Nguyen, Kui Wang, Gang Hu, Xuran Wang, Zhen Miao, Joshua A Azevedo, EunRan Suh, Vivianna M Van Deerlin, David Choi, Kathryn Roeder, Mingyao Li, and Edward B Lee. Apoe and trem2 regulate amyloid-responsive microglia in alzheimer's disease. *Acta Neuropathol*, 140(4):477–493, Oct 2020.

[25] Tomasz J Nowakowski, Aparna Bhaduri, Alex A Pollen, Beatriz Alvarado, Mohammed A Mostajo-Radji, Elizabeth Di Lullo, Maximilian Haeussler, Carmen Sandoval-Espinosa, Siyuan John Liu, Dmitry Velmeshev, et al. Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. *Science*, 358(6368):1318–1323, 2017.

[26] Neelroop N Parikshak, Rui Luo, Alice Zhang, Hyejung Won, Jennifer K Lowe, Vijayendran Chandran, Steve Horvath, and Daniel H Geschwind. Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell*, 155(5):1008–21, Nov 2013.

[27] Damon Polioudakis, Luis de la Torre-Ubieta, Justin Langerman, Andrew G Elkins, Xu Shi, Jason L Stein, Celine K Vuong, Susanne Nichterwitz, Melinda Gevorgian, Carli K Opland, et al. A single-cell transcriptomic atlas of human neocortical development during mid-gestation. *Neuron*, 103(5):785–801, 2019.

[28] Aditya Pratapa, Amogh P Jalihal, Jeffrey N Law, Aditya Bharadwaj, and T M Murali. Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. *Nat Methods*, 17(2):147–154, 02 2020.

[29] John E Reid and Lorenz Wernisch. Pseudotime estimation: deconfounding single cell time series. *Bioinformatics*, 32(19):2973–2980, 2016.

[30] Wouter Saelens, Robrecht Cannoodt, Helena Todorov, and Yvan Saeys. A comparison of single-cell trajectory inference methods. *Nature biotechnology*, 37(5):547–554, 2019.

[31] F Kyle Satterstrom, Jack A Kosmicki, Jiebiao Wang, Michael S Breen, Silvia De Rubeis, Joon-Yong An, Minshi Peng, Ryan Collins, Jakob Grove, Lambertus Klei, et al. Large-scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. *Cell*, 180(3):568–584, 2020.

[32] Kelly Street, Davide Risso, Russell B Fletcher, Diya Das, John Ngai, Nir Yosef, Elizabeth Purdom, and Sandrine Dudoit. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC genomics*, 19(1):477, 2018.

[33] Tim Stuart, Andrew Butler, Paul Hoffman, Christoph Hafemeister, Efthymia Papalexi, William M Mauck, 3rd, Yuhan Hao, Marlon Stoeckius, Peter Smibert, and Rahul Satija. Comprehensive integration of single-cell data. *Cell*, 177(7):1888–1902.e21, 06 2019.

[34] Fran Supek, Matko Bošnjak, Nives Škunca, and Tomislav Šmuc. Revigo summarizes and visualizes long lists of gene ontology terms. *PloS one*, 6(7):e21800, 2011.

[35] Dawn Thompson, Aviv Regev, and Sushmita Roy. Comparative analysis of gene regulatory networks: from network reconstruction to evolution. *Annual review of cell and developmental biology*, 31:399–428, 2015.

[36] Jinjin Tian, Jiebiao Wang, and Kathryn Roeder. Esco: single cell expression simulation incorporating gene co-expression. *Bioinformatics*, 02 2021.

[37] Dmitry Velmeshev, Lucas Schirmer, Diane Jung, Maximilian Haeussler, Yonatan Perez, Simone Mayer, Aparna Bhaduri, Nitasha Goyal, David H Rowitch, and Arnold R Kriegstein. Single-cell genomics identifies cell type–specific molecular changes in autism. *Science*, 364(6441):685–689, 2019.

[38] A Jeremy Willsey, Stephan J Sanders, Mingfeng Li, Shan Dong, Andrew T Tebbenkamp, Rebecca A Muhle, Steven K Reilly, Leon Lin, Sofia Fertuzinhos, Jeremy A Miller, Michael T Murtha, Candace Bichsel, Wei Niu, Justin Cotney, A Gulhan Ercan-Sencicek, Jake Gockley, Abha R Gupta, Wenqi Han, Xin He, Ellen J Hoffman, Lambertus Klei, Jing Lei, Wenzhong Liu, Li Liu, Cong Lu, Xuming Xu, Ying Zhu, Shrikant M Mane, Ed S Lein, Liping Wei, James P Noonan, Kathryn Roeder, Bernie Devlin, Nenad Sestan, and Matthew W State. Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell*, 155(5):997–1007, Nov 2013.

[39] Shixiong Zhang, Xiangtao Li, Qiuzhen Lin, and Ka-Chun Wong. Review of single-cell rna-seq data clustering for cell type identification and characterization. *arXiv preprint arXiv:2001.01006*, 2020.

[40] Juan Zhao, Yiwei Zhou, Xiujun Zhang, and Luonan Chen. Part mutual information for quantifying direct associations in networks. *Proc Natl Acad Sci U S A*, 113(18):5130–5, May 2016.

[41] Lingxue Zhu, Jing Lei, Bernie Devlin, and Kathryn Roeder. Testing high-dimensional covariance matrices, with application to detecting schizophrenia risk genes. *The annals of applied statistics*, 11(3):1810, 2017.

[42] Lingxue Zhu, Jing Lei, Lambertus Klei, Bernie Devlin, and Kathryn Roeder. Semisoft clustering of single-cell data. *Proceedings of the National Academy of Sciences*, 116(2):466–471, 2019.

Additional Files

Additional file 1 — Supplementary Information

PDF file of supplementary information.

Additional file 2 — Table S13

Genes in each cluster of trajectory from Developing Cortex Atlas dataset.

Additional file 3 — Table S14

GO terms of the stable genes in each cluster of trajectory from Developing Cortex Atlas dataset.

Additional file 4 — Table S15

Table of leverage genes and top 90% differential network genes from ASD Brain dataset.